

Dokumente, Zugang, Präsentation

Langzeiterhaltung digitaler Inhalte für das CultLib-Projekt II

Wir sichern Kulturgut in digitaler Form
und machen es frei verfügbar,
heute und für zukünftige Generationen.

CultLib-ID: 5b8cc70b-6540-42ff-97d7-a15ee49eec78

Version: 1.0

Datum: 11. April 2014

Verfasser: Dr. sc. math. Hartwig Thomas



<http://creativecommons.org/licenses/by-sa/3.0/ch/>

Übersicht

Das Projekt [CultLib](#) des Vereins [Digitale Allmend](#) bezweckt die Gründung einer schweizerischen Stiftung *Pro Cultura Libera* welche ein digitales Repository freier, kultureller Werke aufbaut und betreibt.

Die Berichte über Aspekte der Langzeiterhaltung¹ erklären, wie CultLib technisch konzipiert ist und dienen als Anforderungen an die Umsetzung. Ausserdem können sie für Entscheidungsträger und IT-Architekten von Memo-Institutionen von Interesse sein.

In diesem ersten Bericht über Erhaltungsstrategien wird dargelegt, dass sich die Langzeiterhaltung digitaler Inhalte nicht in ihrer blossen Erhaltung erschöpft. Auch ihr Zugang muss langfristig gesichert sein und sie müssen auf verschiedenste Weise präsentiert werden.

Für die Dokumente, den Zugang und die Präsentation spielen unterschiedliche Zeithorizonte eine Rolle.

¹ Der Begriff Nachhaltigkeit wird im Zusammenhang mit der Langzeiterhaltung bewusst vermieden, weil die virtuelle Allmend in ihrer Ausdehnung nicht beschränkt ist, und deshalb die mit dem Begriff Sustainability adressierte Übernutzung, die traditionelle Tragödie der Allmend, im Bereich der digitalen Inhalte keine Gefahr darstellt. Das Ziel der Tradierung kultureller Inhalte ist würdig und gewichtig genug und hat nicht nötig, sich auf die Leere übernutzter Warmluft-Begriffe zu stützen.

Inhaltsverzeichnis

1	Wie lang ist Langzeit?.....	4
1.1	Dokumenterhaltung: langfristig (50-200 Jahre).....	4
1.1.1	Katalogsystem: 5-10 Jahre.....	5
1.2	Stabiler Zugang über das Web: mittelfristig (20-50 Jahre).....	5
1.2.1	Webanwendung: 5-10 Jahre.....	6
1.3	Präsentation und Design: kurzfristig (5-10 Jahre).....	6
2	Fazit.....	8

1 Wie lang ist Langzeit?

CultLib ist ein Digitales Museum, eine Digitale Bibliothek, ein Digitales Archiv. Als Memo-Institution setzt sich CultLib zum Ziel, digitale Inhalte für zukünftige Generationen aufzubewahren. Die Aufgabe einer Memo-Institution beschränkt sich aber nicht auf die Erhaltung. Fast ebenso wichtig ist die Vermittlung der vor dem Vergessen bewahrten Inhalte, welche dem interessierten Publikum den Zugang zu ihnen erschliesst.

Für ein digitales Archiv² erfolgt der Zugang heute sinnvollerweise über das Internet. Noch genauer: über das, durch das HTTP-Protokoll definierte, World Wide Web. Dabei ist daran zu denken, dass immer öfter nicht nur die direkte Vermittlung an menschliche Nutzer sondern auch die indirekte Vermittlung eine Rolle spielen wird, wo Programme an Stelle von Menschen auf die Inhalte zugreifen und sie in irgendeiner Weise für die menschlichen Endnutzer aufbereiten. Ein heute weit verbreitetes Beispiel sind Roboter von Google und ähnlichen Institutionen, welche die Inhalte durchforsten, indexieren und für die Suche zugänglich machen.

Schliesslich spielt die eingängige Präsentation der Inhalte nicht nur bei traditionellen analogen Museen, sondern auch bei digitalen Archiven eine wichtige Rolle und bestimmt den Erfolg einer Sammlung ganz wesentlich.

Nun ist ziemlich klar, dass für Präsentation, Zugang und Dokumente ganz verschiedene Zeithorizonte relevant sind. Selbst wenn ein normalisiertes Dokument in fünfzig bis zweihundert Jahren noch entziffert und verstanden werden kann, dürfte eine heutige Website schon in zehn Jahren nicht mehr funktionsfähig oder mindestens hoffnungslos antiquiert und für moderne Geräte unzugänglich sein.

Im Folgenden wird der Zeithorizont für diese Schichten des CultLib-Archivs näher ins Auge gefasst und die einzusetzende Technologie unter diesem Aspekt beleuchtet. Als einzigen Anhaltspunkt für die Entwicklung der Zukunft dient uns die Entwicklung in der Vergangenheit und ihre Geschwindigkeit. Es ist allerdings anzunehmen, dass die Geschwindigkeit der Entwicklung in Zukunft eher abnimmt, je weiter die Technologien und Infrastrukturen verbreitet sind, soweit von einer „Geschwindigkeit der Entwicklung“ überhaupt sinnvoll die Rede sein kann.

Alle hier aufgeführten Überlegungen zur digitalen Erhaltung basieren auf der Annahme, dass die Menschheit nicht wieder auf ein nicht-digitales Zeitalter zurückfällt. Ein solcher Rückfall ist nicht grundsätzlich auszuschliessen. Sofern er nicht sehr plötzlich erfolgt, können eventuell vorher analoge „Printouts“ von Dokumenten hergestellt werden. In einer nicht-digitalen Zukunft stellt sich natürlich die Frage nach der Erhaltung digitaler Inhalte nicht mehr.

1.1 Dokumenterhaltung: langfristig (50-200 Jahre)

Wenn Dokumente in sorgfältig ausgewählten Normalformaten gespeichert werden, wie dies im ersten Artikel dieser Serie ausgeführt wird, ist es wahr-

² Wo im Folgenden von Archiven die Rede ist, sind ganz allgemein Memo-Institutionen wie Bibliotheken, Museen und eigentliche Archive gemeint; das Wort Archivierung wird synonym für Erhaltung oder Aufbewahrung benutzt.

scheinlich, dass diese in den nächsten hundert Jahren in keiner Weise verändert werden müssen. Ja es ist sogar erwünscht, dass sie unangetastet bleiben und nicht modernisiert werden, weil sonst die Authentizität nicht mehr gewährleistet werden kann. Es ist auch denkbar, dass solche Dokumente länger als hundert Jahre unverändert aufbewahrt werden können. Andererseits hat sich die Informationstechnologie in den letzten 50 Jahren derart rasant entwickelt, dass es vorstellbar scheint, dass schon in weniger als hundert Jahren das Konzept von endlichen Byte-Strömen und Dateiformaten nicht mehr anwendbar ist. Um aber ernsthaft von Langzeiterhaltung zu reden, sollte mindestens ein Zeithorizont von fünfzig Jahren mit höchster Wahrscheinlichkeit garantiert sein.

1.1.1 Katalogsystem: 5-10 Jahre

Als zentrales Werkzeug im Archiv dient der Katalog oder das Archivinformationssystem (AIS), wie es im modernen Archivstandard heisst. Ein solcher Katalog ist normalerweise in einer relationalen Datenbank gespeichert. Datenbankinhalte haben erwartungsgemäss eine mittels Migration garantierte Lebensdauer von 10-30 Jahren. Die konkret eingesetzte Datenbanktechnologie kann schwerlich für länger als 10 Jahre garantiert werden. Ausserdem ist weltweit mit dem Entstehen eines Geflechts verteilter Archive zu rechnen, wo das Konzept einer monolithischen Datenbank eines einzelnen Archivs ohnehin in Frage gestellt wird.

Dies illustriert die Wichtigkeit eingebetteter Metadaten im Archivgut. Anhand dieser Daten kann eine Katalogdatenbank – oder später ein verteilter Katalog vieler Institutionen – jederzeit aus den Dokumenten neu erzeugt werden. Die kürzere Lebensdauer des Katalogs beeinflusst die längere Lebensdauer der Dokumente nicht.

1.2 Stabiler Zugang über das Web: mittelfristig (20-50 Jahre)

Die Zugänglichkeit von Inhalten für Menschen und Programme wird durch eine Schnittstelle definiert. Eine solche Kommunikationsschnittstelle wird auch „Protokoll“ genannt. Etwas genauer handelt es sich sogar um einen Protokollstapel, der – stark vereinfacht – aus drei hier relevanten Schichten besteht:

- Internet (IP-Protokoll)
- World Wide Web (Domänensystem, HTTP, URLs, XML)
- CultLib REST-Protokoll

Das eigentliche Internet-Protokoll (IP) wird sich zwar weiterentwickeln (zum Beispiel wird aktuell gerade sukzessive von IPv4 auf IPv6 umgestellt), aber kaum in den nächsten 50 Jahren grundlegend verändert werden. Die technische und institutionelle Infrastruktur des Internet ist inzwischen so stark verankert, dass seine völlige Ablösung in den nächsten 50 Jahren unwahrscheinlich scheint. Selbst wenn neben dem Internet eine neue Kommunikationstechnologie entstehen sollte, von der heute allerdings keine Ansätze sichtbar sind, so wird das Internet daneben noch viele Jahre fortbestehen, wie etwa die traditionelle Leitungsvermittlung der Festnetztelefonie 40 Jahre lang neben dem

Internet weiterbestand und erst heute schrittweise durch die Internet-Telefonie abgelöst wird.

Die Kollektion von Protokollen, welche das World Wide Web ausmachen, basiert zentral auf dem Hypertext Transfer Protokoll (HTTP). Dieses wurde 1990 konzipiert. 1999 entstand die Version 1.1. Diese hat sich bis heute als stabil erwiesen. Eine Version 2.0, die es ablöst, ist momentan in Arbeit, wird aber wohl nur rückwärtskompatible Erweiterungen bringen. Es ist zwar wahrscheinlich, dass sich einzelne Aspekte besonders im Bereich Sicherheit und Datenschutz weiterentwickeln. Aber ein ernsthaftes Verschwinden des HTTP-Protokolls, auf welchem heute Milliarden von Webseiten beruhen, ist in den nächsten Jahrzehnten nicht zu befürchten.

Schliesslich wird CultLib ein eigenes Protokoll für den Zugang zu seinen Inhalten definieren müssen. Dieses wird versuchen, die Anforderungen an ein verteiltes Archiv schon zu berücksichtigen. Gemäss dem sogenannten REST-Konzept (Representational State Transfer) wird es vollständig auf dem HTTP-Protokoll basieren. URLs (Uniform Resource Locators) werden persistent und eindeutig immer auf die gleichen Inhalte verweisen. Auf Inhalte wird im XML-Format (eXtensible Markup Language) oder in denormalisierten Dokumentformaten zugegriffen.

Dieses Protokoll ist auf eine Mindestdauer von 20 Jahren auszulegen. Es ist davon auszugehen, dass es rückwärtskompatibel weiterentwickelt wird. Idealerweise kann eine solche Protokollschicht wohl maximal 50 Jahre dienen, sofern sie genügend Verbreitung und Anwendung findet.

1.2.1 Webanwendung: 5-10 Jahre

Eine Webanwendung, welche sich auf die oben erwähnte Katalogdatenbank stützt und das CultLib-Protokoll implementiert, wird wohl kaum länger als fünf bis zehn Jahre überleben, bevor sie grundlegend überholt werden muss. In der Zeit kommen neue Programmiersprachen, neue parallele Hardware auf den Markt und alten Betriebssystemen und Programmiersprachen wird der Support notfalls mit Gewalt entzogen, wie dies heute mit Windows XP der Fall ist.

Die kürzere Lebensdauer der Webanwendung stört aber die längere Geltung des Protokolls nicht.

1.3 Präsentation und Design: kurzfristig (5-10 Jahre)

Letztlich ist es schliesslich die konkrete Webseite und deren Design, welche dem Nutzer die Inhalte präsentiert. Dabei ist zu bedenken, dass sich verschiedenste Anwendungen an derselben Webschnittstelle bedienen und deren Inhalte auf ihre Art präsentieren können.

So ist etwa denkbar, dass Einzelsammlungen in speziell als „Führung“ durch das Archivgut gestaltete Websites erschlossen werden. Verschiedene Gerätetypen (PC, Tablet, Smartphone, ...) benötigen unterschiedliche Gestaltung der Präsentation. Es ist wünschenswert, dass Präsentationen in vielen Sprachen der Welt entstehen.

Auch wenn heute traditionelle Memo-Institutionen mehrheitlich krampfhaft darauf bestehen, dass man Archivgut nur von ihre hauseigenen Servern abru-

fen darf, führt dieser Ansatz unweigerlich in die Isolation. Gerade die Fans einer Institution und ihrer Inhalte werden durch eine solche Abweisung vor den Kopf gestossen. Nur durch Öffnung gegenüber vielfältigen Präsentationen kann das Ziel der breiten Vermittlung kultureller Inhalte erreicht werden.

Externe Präsentationen jeder Art werden von CultLib begrüsst.

Und dann ist Design auch eine Modesache. Nichts veraltet schneller als der gute Geschmack. CultLib wird einen eigenen Design anbieten müssen, weil anfänglich kaum eine externe Präsentation von sich aus entstehen wird. Mit diesem wird eine Zeitlosigkeit angestrebt, wie sie etwa die Google-Suchseite erreichte. Auch dann ist davon auszugehen, dass ein Design der Präsentation höchstens zehn Jahre unverändert einsetzbar sein wird. Besonders in der Anfangszeit wird sich auch das Protokoll und der abgedeckte Bereich noch stark entwickeln.

Schliesslich sind auch die Technologien des Design sehr kurzlebig. Heute werden wir clientseitig HTML5, JavaScript mit AJAX einsetzen, welche in allen gängigen Browsern unterstützt werden. Proprietäre Technologien und Programmiersprachen vermeiden wir dabei so gut es geht. Trotzdem ist damit zu rechnen, dass eine heutige Implementation in fünf Jahren obsolet ist.

2 Fazit

Obwohl normalisierte Dokumente während langen Zeiträumen als endliche Byte-Sequenzen unverändert erhalten bleiben, benötigt ein digitales Archiv laufende Pflege nicht nur der Aufnahme neuer Inhalte sondern auch zur Aufrechterhaltung der Zugänglichkeit und der Vermittlung. Dabei sind die unterschiedlichen Zeithorizonte für die verschiedenen Ebenen zu beachten.